



Overview of the EDRN Informatomics Center: eDRN Catalog and Archive Service

Chris A. Mattmann

Schedule

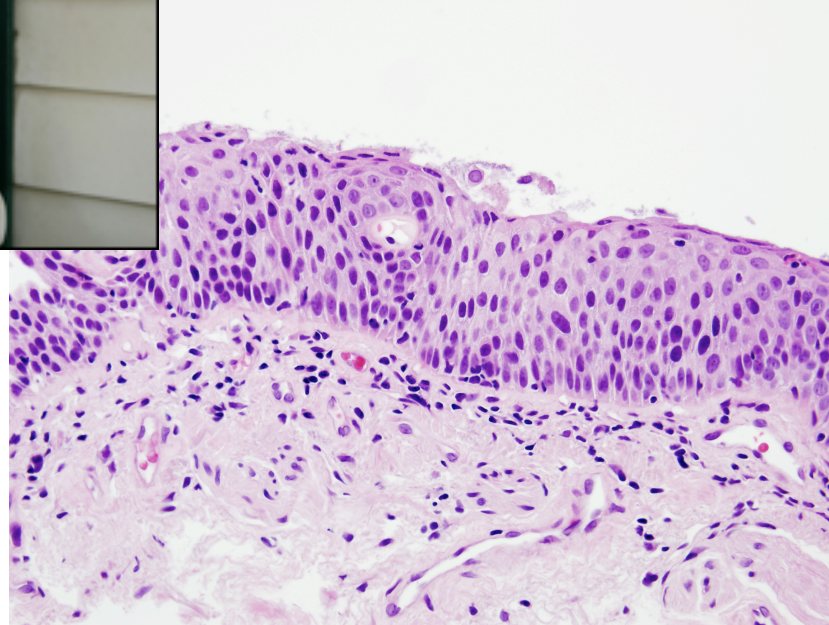
- What is the EDRN IC?
- Who's involved?
- What are the types of applications we build?
- How do we build them?
- Focus: eCAS
- Results to date
- Where we're headed

The Early Detection Research Network

- EDRN is a network of 40+ institutions all performing research geared towards the discovery of cancer *biomarkers*, which are early indicators of onset of disease
- NCI/NIH funded program
 - 100s of millions of dollars of investment
 - Recently renewed after NIH Board of Scientific Advisors advised NIH of the strategic value in EDRN and the existing successes

Cancer Biomarkers

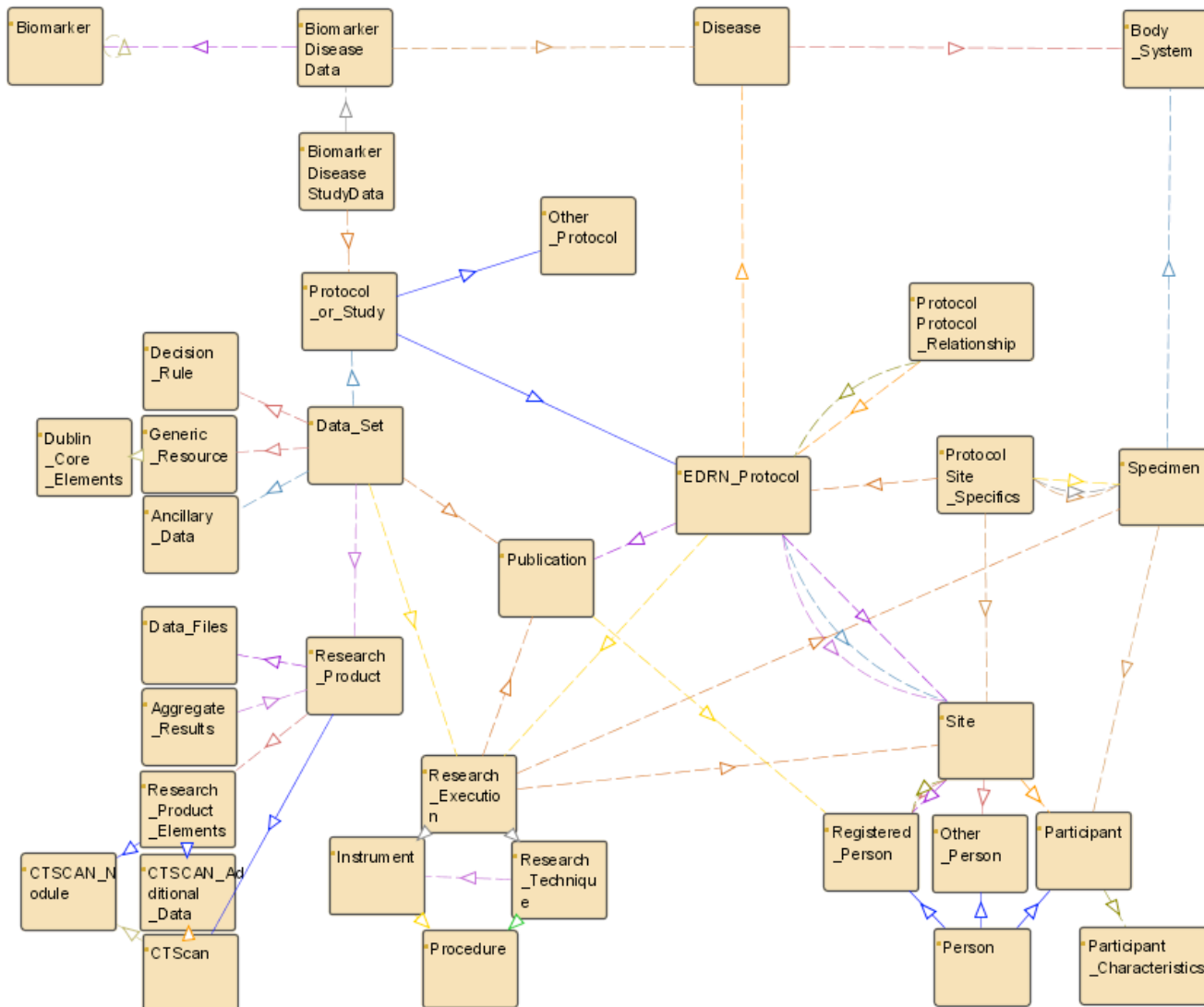
- Indicators of early onset of disease
 - Important to capture information about:
 - Characteristics of patients and onsets (epidemiological information)
 - Causes



Types of information important in EDRN

- Specimen Inventories
 - At each site, maps specimens collected (blood, sputum, etc.) to patient characteristics
- Studies and Protocols
 - Information about studies conducted in the EDRN and results (publications, outputs)
- Biomarkers
 - Information about indicators of early disease
- Science Data
 - Outputs of experiments on specimens, regarding biomarkers, driven by particular studies and protocols

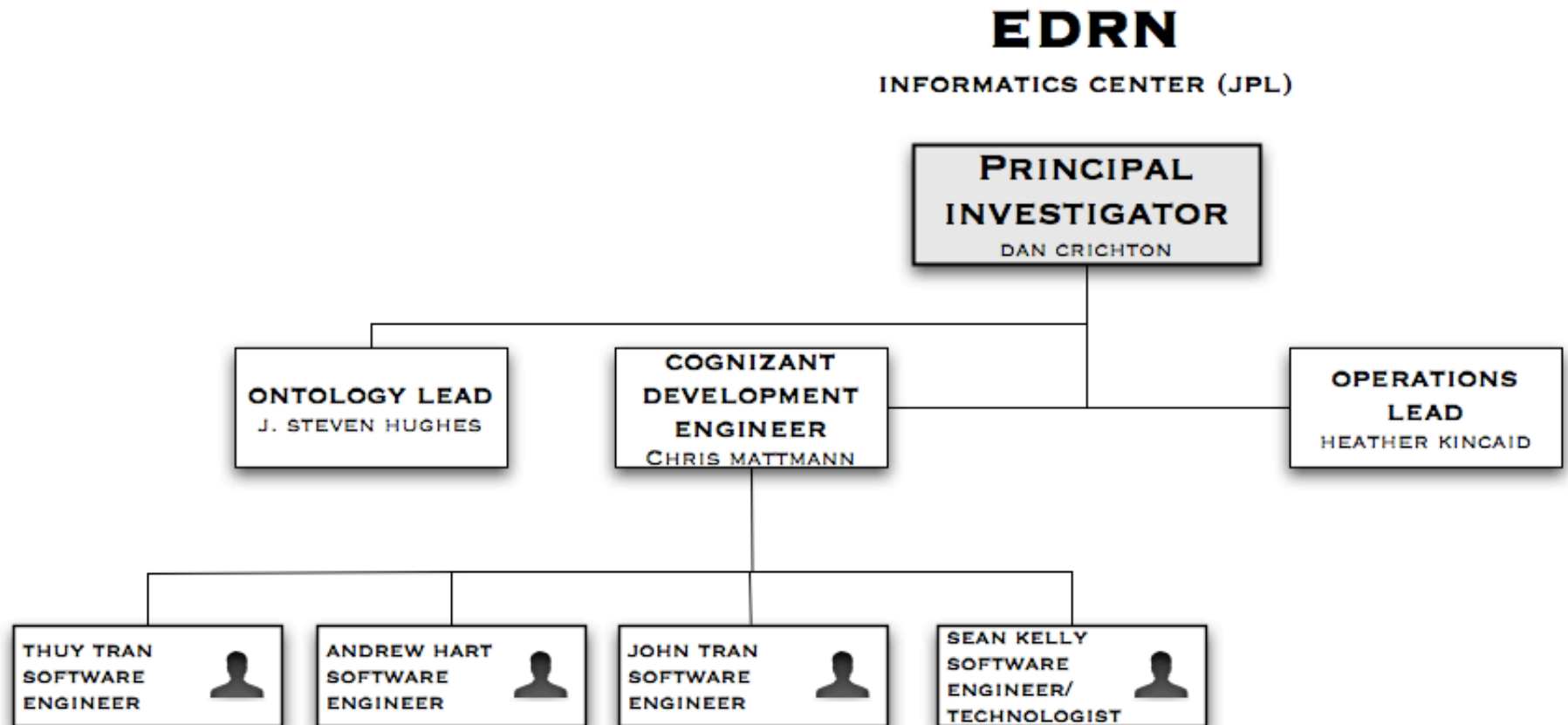
Types of information important in EDRN



The EDRN Informatics Center's Role

- The EDRN Informatics Center develops software solutions to support the Early Detection Research Network's research of cancer biomarkers and development of cancer-fighting tools.
 - Develop software solutions and provide informatics expertise to assist EDRN centers and NCI in capturing and disseminating important EDRN information mentioned before

Informatics Center Organization



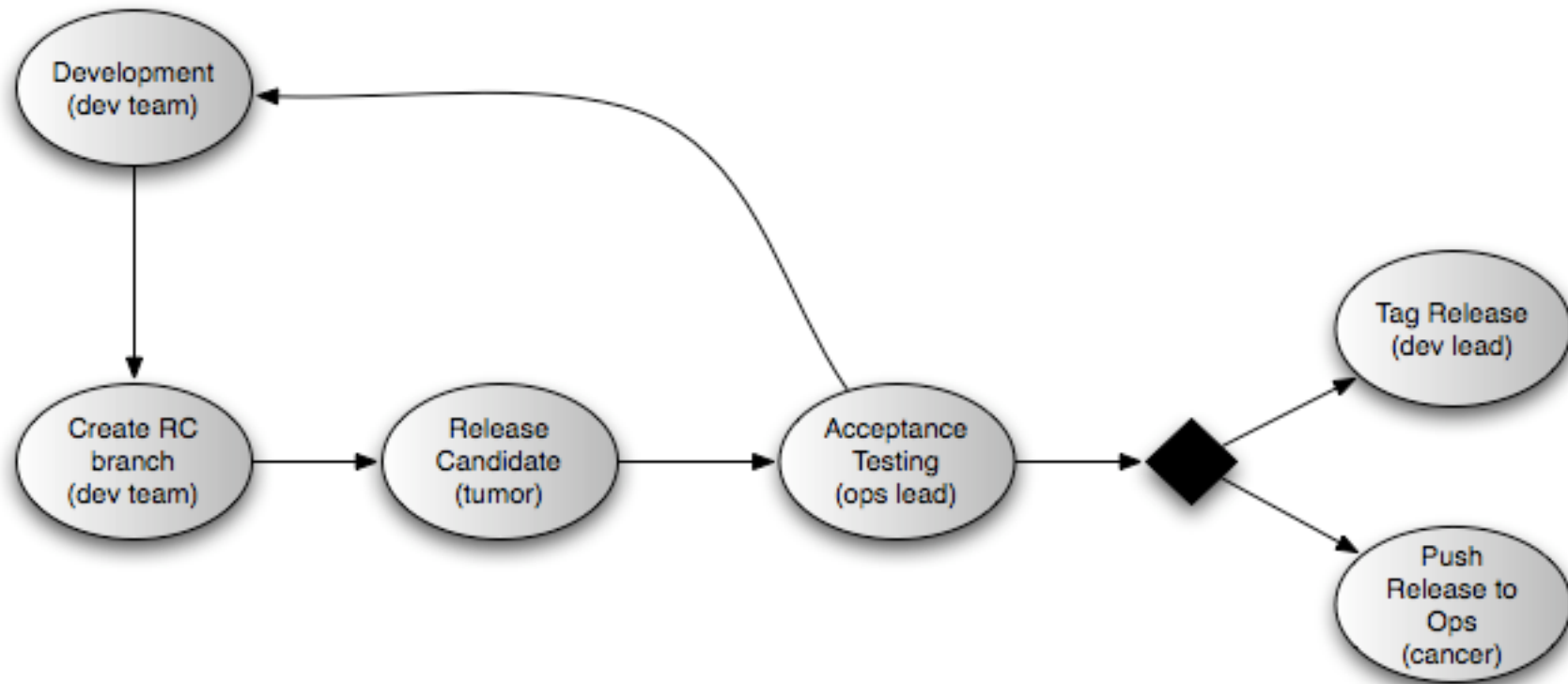
EDRN Informatics Center Hardware

- Tumor
 - FusionX5 1U High-performance Node w/ 1x 5140 Dual-core CPUs
 - 0.5 TB total space
 - Development machine
- Cancer
 - FusionX5 1U High-performance Node w/ 2x 5140 Dual-core CPUs
 - 1 TB total space
 - Operations machine



EDRN Informatics Center: Development Process

- *Read: agile*



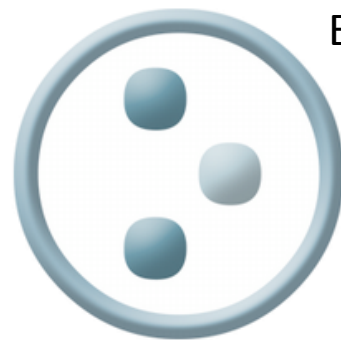
Use of modern software process tools



Issue tracking



Source code
repository



Plone

EDRN IC site



Repository
browsing



Mailing lists



Acceptance
testing

Let's talk about eCAS

- Data Warehousing Application
 - Store raw and processed data from local laboratories
 - NCI eCAS that holds curated “high quality” eCAS data published from local sites
- Many potential applications
 - Specimen images, datasets, instrument generated products, etc.

eCAS High Level Architecture



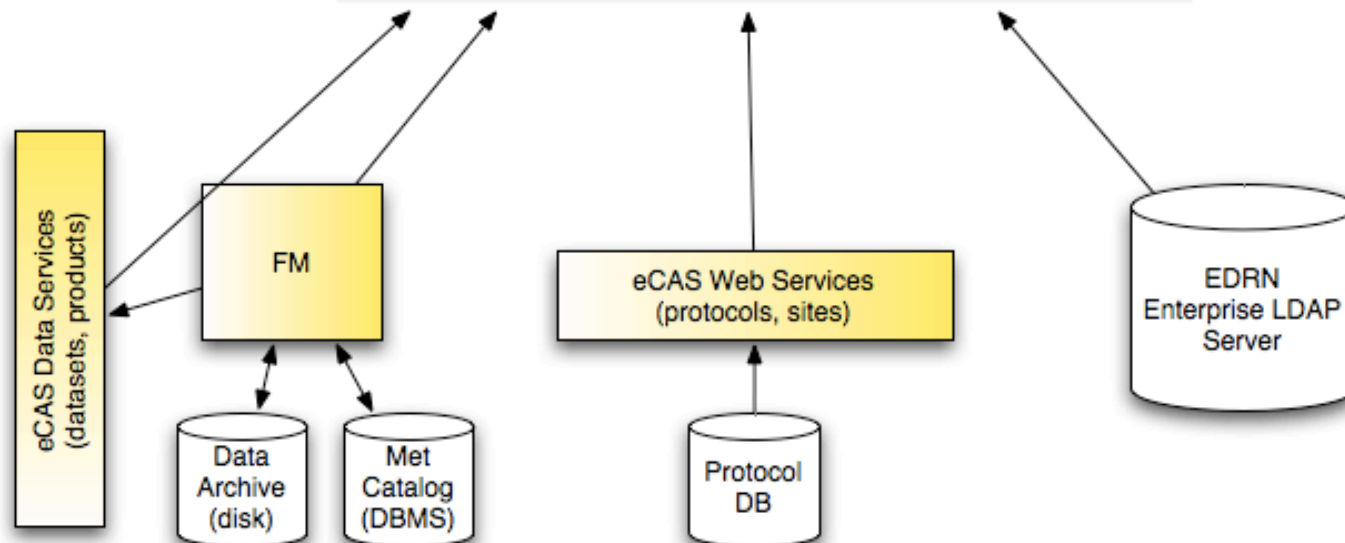
Available Datasets

- Clinical Utility of Urinary CD90 as a Biomarker for Prostate Cancer Detection

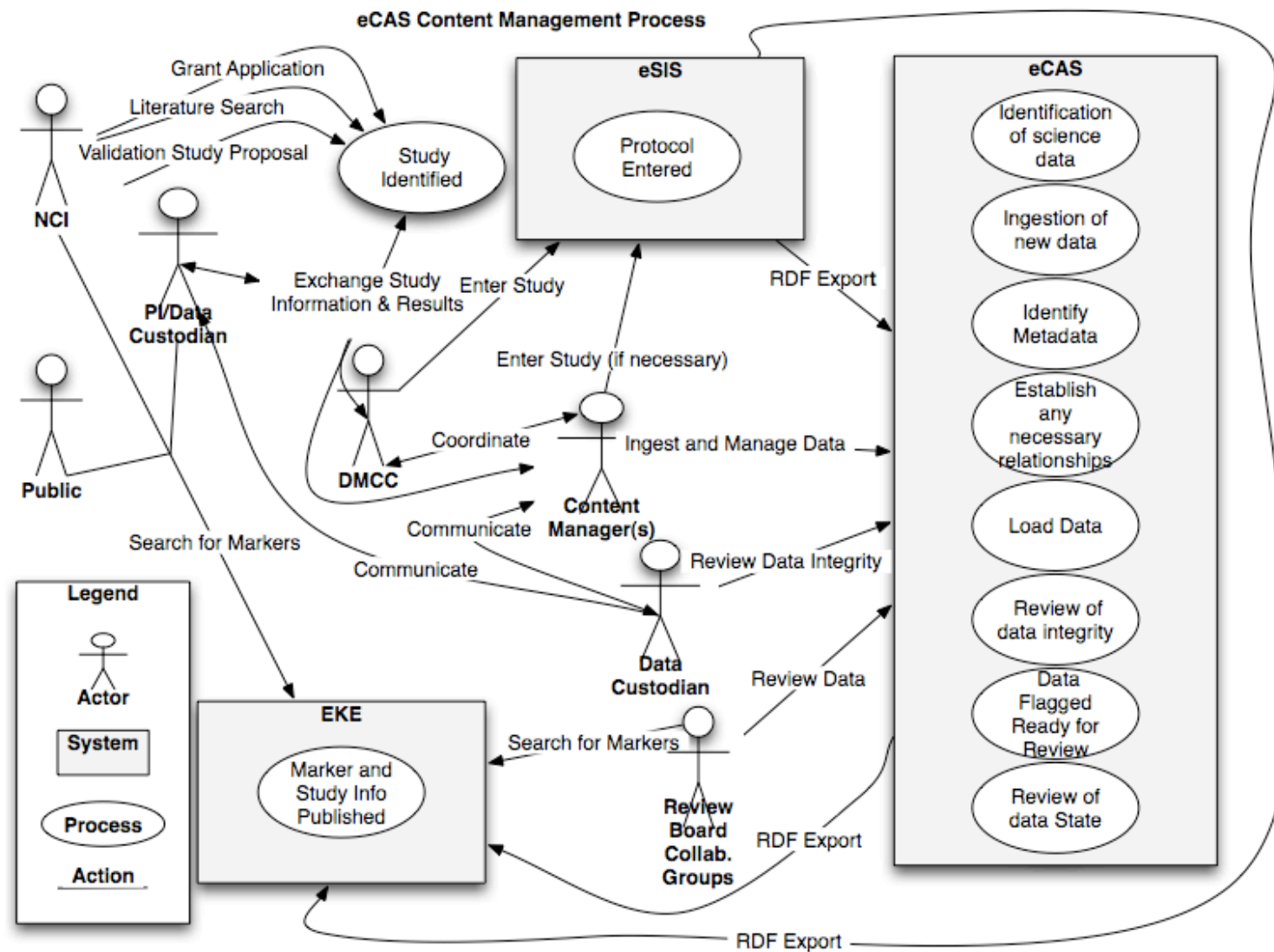
University of Washington Immunohistochemistry data. ([4 products](#))
[PI: Alvin Liu, Organ: Prostate, Collaborative Group: Prostate and Urologic]

- EDRN-WHI Pre-Clinical Colon Ca Specimens

University of Pittsburgh Cancer Institute WHI Mass Spectrometry ([13 products](#))
[PI: William Bigbee, Organ: Colon, Collaborative Group: GI and Other Associated]



eCAS Curation Process

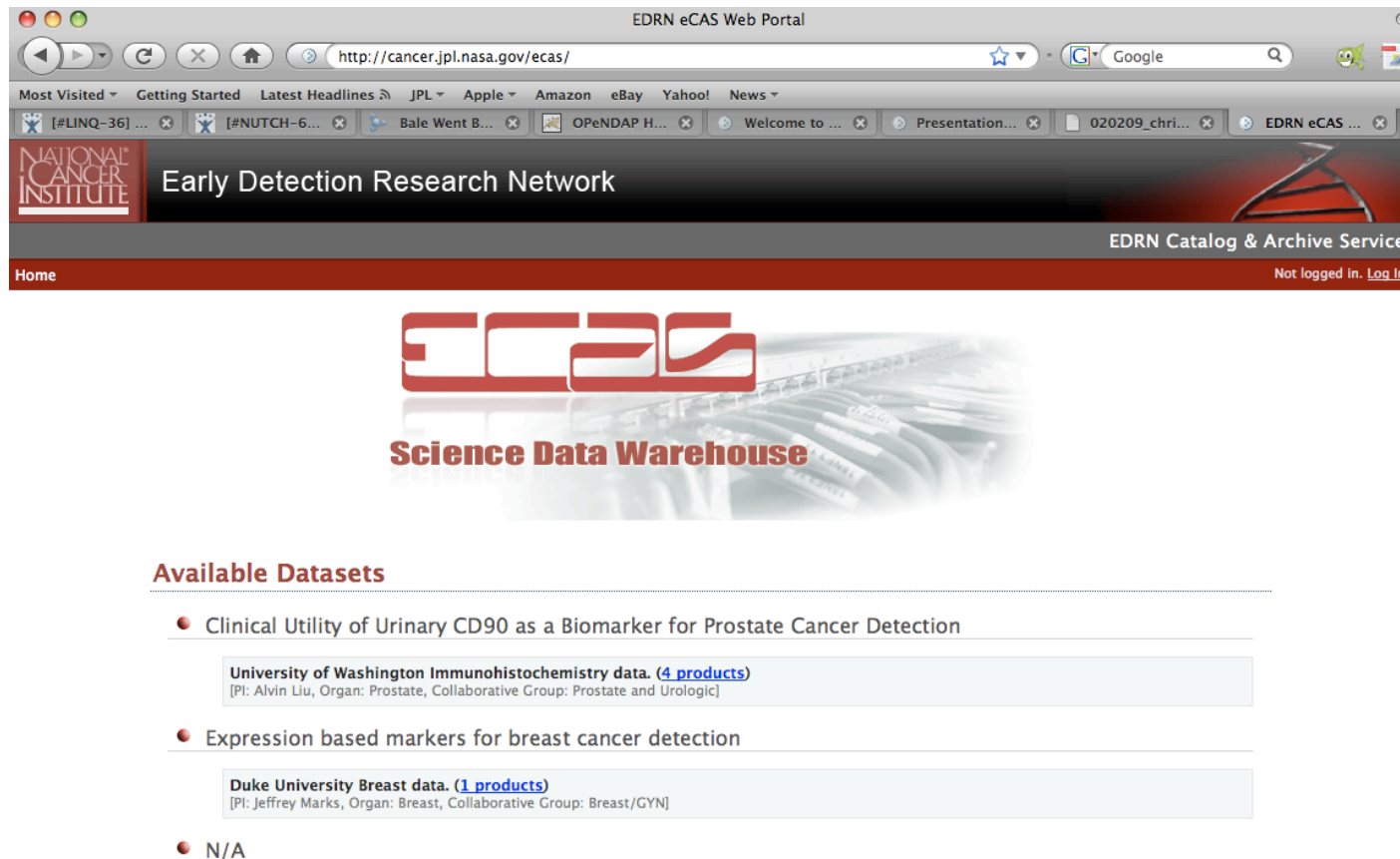


Role of DMCC and NCI

- DMCC
 - Help to derive metadata standards for different dataset types
 - Annotation of datasets
 - Delivery of data (processed)
- NCI
 - Encourage PIs to submit data to eCAS
 - Encourage PIs to run local laboratory eCAS
 - Capture Prostate 2000 data along with disclaimer

eCAS Interface: Splash Page

- <http://ecas.jpl.nasa.gov>
 - Operational 0.4.2 beta release



The screenshot shows a web browser window titled "EDRN eCAS Web Portal" with the URL "http://cancer.jpl.nasa.gov/ecas/". The browser's address bar and search bar are visible. Below the browser window, the splash page features the National Cancer Institute logo and the text "Early Detection Research Network". A red banner at the top right says "EDRN Catalog & Archive Service" and "Not logged in. Log In". The main content area has a large red "eCAS" logo and the text "Science Data Warehouse" over a background image of server racks. Below this, the "Available Datasets" section lists three items:

- Clinical Utility of Urinary CD90 as a Biomarker for Prostate Cancer Detection
 - University of Washington Immunohistochemistry data. ([4 products](#))
[PI: Alvin Liu, Organ: Prostate, Collaborative Group: Prostate and Urologic]
- Expression based markers for breast cancer detection
 - Duke University Breast data. ([1 products](#))
[PI: Jeffrey Marks, Organ: Breast, Collaborative Group: Breast/GYN]
- N/A

eCAS Interface: Logged in Dataset View

The screenshot shows a web browser window titled "EDRN ecas user interface". The address bar displays the URL "http://cancer.jpl.nasa.gov/ecas/dataset.php?typeID=urn:edrn:UABPreinva...". The browser's toolbar includes navigation buttons and a search bar. Below the browser window, the page header features the "NATIONAL CANCER INSTITUTE" logo and the text "Early Detection Research Network". A navigation bar includes links for "Home", "Protocol Dataset: UABPreinvasiveNeoplasiaUnProcessedImages", and a login status "Logged in as mattmann. Log Out".

UABPreinvasiveNeoplasiaUnProcessedImages

Dataset Metadata: [less information \[-\]](#)

Abstract	APC signaling pathway data including images and powerpoints related to carcinoma. Test image and summary data contributed by Bill Grizzle et al.
DataCustodian	Bill Grizzle
Siteld	University of Alabama at Birmingham (Biomarker Reference Laboratories)
Author	Bill Grizzle
ProtocolId	N/A
Title	UAB Test Preinvasive Neoplasia Image and Summary Data
Date	2008-12-16T18:24:00.000Z
DatasetId	urn:edrn:UABPreinvasiveNeoplasiaUnProcessedImages
OrganId	Cervix
CollaborativeGroup	Breast/GYN

Download Dataset as a Zip Archive: [less information \[-\]](#)

Download Zip Archive:	Click Here
-----------------------	----------------------------

5 Products Associated With This Dataset:

- [cx-2.ppt](#)
- [cx-1.ppt](#)
- [leuk 1.ppt](#)
- [APC PATH.ppt](#)
- [colon PATH.ppt](#)

eCAS Interface: Logged in Product View

The screenshot shows a web browser window titled "EDRN ecas user interface". The address bar displays the URL: <http://cancer.jpl.nasa.gov/ecas/product.php?productID=4e376d26-cd91-11dd-99fb-f72e743664d5>. The browser's toolbar includes navigation buttons and a search bar with the text "Google". Below the browser window, the page header features the National Cancer Institute logo and the text "Early Detection Research Network". A secondary header reads "EDRN Catalog & Archive Service". The main content area has a red navigation bar with the text "Home / UABPreinvasiveNeoplasiaUnProcessedImages / Product: leuk 1.ppt" and a login status "Logged in as mattmann. Log Out". The main heading is "UABPreinvasiveNeoplasiaUnProcessedImages / leuk 1.ppt". The page is divided into two columns. The left column contains the "Product Name: leuk 1.ppt" with a "more information [+]" link, and "Product Metadata: 4e376d26-cd91-11dd-99fb-f72e743664d5" with a "less information [-]" link. Below the metadata is a table with the following data:

FileLocation	/data/ingest/grizzle/preinvasive_neoplasia
CAS.ProductName	leuk 1.ppt
SiteShortName	ualb
CAS.ProductReceivedTime	2008-12-18T21:53:05.719-08:00
SpecimenId	Unknown
InstrumentId	immuno
Filename	leuk 1.ppt
ProductType	UABPreinvasiveNeoplasiaUnProcessedImages

The right column is titled "Files Associated With This Product:" and lists the file "leuk 1.ppt" with a size of "3.8 MB" and a "Mime Type:" field. A "save" link is provided below the mime type.

Capabilities

- Current (as of 0.4.2 beta)
 - Download datasets as zip files
 - Zips include zips of products (and their metadata), as well as dataset-level metadata
 - Download products as zip files
 - Browse of dataset and product metadata
 - Dump RDF dataset and product information into EKE
 - Pagination of results
- Planned (in next 2 months, 2-4 releases)
 - Curation interface
 - Talk more about this later, with John's slides
 - Bug fixes/maintenance

Current Datasets

- SELDI Phase I and II, MALDI Dilution Data, Miscellaneous PI data, (Immunohistochemistry, LabMAP Mass Spectrometry, FISH MSA data), WHI datasets
- Test data from Bill Grizzle (as of 0.3.0 beta release)
 - EGFR Translocation Data
 - Pre-invasive Neoplasia Image and Summary Data
- Sites
 - NIST, FHCRC, EVMS, UAB, Duke, Univ Pitt, Univ Wash, Wayne State University

Security


- Current
 - Integration with EDRN Informatics Center LDAP server
 - Basic authentication, to guard datasets and products, and download
- Planned
 - Granular authorization, based on roles
 - Need to develop dataset, product-level policy, as well as metadata attribute browse, and downloading policy

Release Policies

- Software
 - Documented at:
<https://cancer.jpl.nasa.gov/documents/applications/catalog-and-archive-service/deploying-the-ecas-to-operations>
 - Describes how to deploy the eCAS software to operations
- Datasets
 - As datasets are acquired from PIs (via mobile eCAS, or via WebDAV, or via CD/DVD), there is a vetting process involving curation activities
 - Dataset-level Metadata curation
 - Metadata extraction for Product-level curation
 - Bulk Ingest, or product-by-product
 - Validation/Verification
 - Performed on tumor (dev) first, then performed on cancer (ops)
 - Datasets released currently along with software, but can be a separate activity as well

Data Sharing

- Defined in part by security policies and authorization at the eCAS/eCAS curation level
- Defined also in part through RDF dump of product/dataset metadata to portal and ability to download datasets
- Requests for eCAS data from private and research communities
 - Data Disclaimer
- eCAS meant to be a software/data enabler, not a lawyer



Early Detection Research Network

Early Detection Research Network

[Home](#) / [Biomarkers](#)

Biomarker Directory:


Browse the directory listing, or search by Title using the box on the right.


Displaying Results: 1-15 of 20


1


2


Name
DCP
CA125
KLK-6
CA72-4
B7-H4
HE4
CA 19-9
CA 15-3
Nephropontin
SMRP
MMP9
CRTL1
Microstallia Inertskilley Agg...



Calendar



Division of Cancer Prevention


Cancer Biomarkers Research Group


Find EDN Protocols


Funding Opportunities


EDN Bookshelf



Need Help?
Contact us by phone,
Web, and email
1-800-4-CANCER

Early Detection Research Network








[Home](#)
[About EDNR](#)
[EDNR Investigators](#)
[Committees](#)
[Organ Groups](#)
[Resources](#)
[Publications](#)

Link: [Return to Map View](#)

[expand \[+\]](#)

[hide \[-\]](#)

Search Results

[15453 MC-3F13.jpg](#)
 Pathology Score: 1100
 Participant Id: [11652](#)
 Source: Univ. Colorado

[15453 MC-3F13.jpg](#)
 Pathology Score: 1100
 Participant Id: [11652](#)
 Source: Univ. Colorado

[15453 MC-3F13.jpg](#)
 Pathology Score: 1100
 Participant Id: [11652](#)
 Source: Univ. Colorado

[15453 MC-3F13.jpg](#)
 Pathology Score: 1100
 Participant Id: [11652](#)
 Source: Univ. Colorado

[15453 MC-3F13.jpg](#)
 Pathology Score: 1100
 Participant Id: [11652](#)
 Source: Univ. Colorado

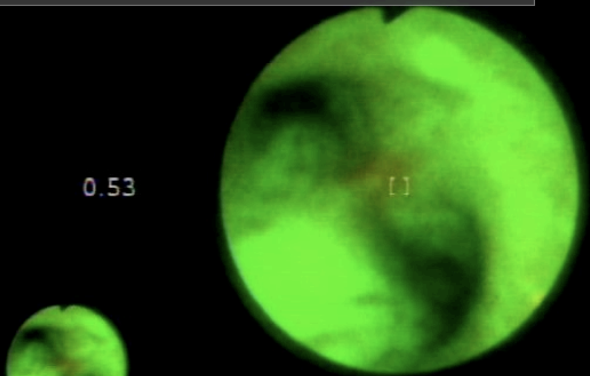
[15453 MC-3F13.jpg](#)
 Pathology Score: 1100
 Participant Id: [11652](#)
 Source: Univ. Colorado

[15453 MC-3F13.jpg](#)
 Pathology Score: 1100
 Participant Id: [11652](#)
 Source: Univ. Colorado

Identifier: 53387ae8-5cd6-11dd-a04c-0b62437fde8b

Title: 15453 MC-3F13.jpg Specimen Collected Date: 07282008

Study Participant Id: 11652 Bronchial Site Code: RB3A/B Pathology Score: 1100



[Scie](#)

[Coll](#)

How to

Welcome
(NCI), bri
applicati

Biomarker Search: Lung

Patient Characteristics

Patient ID:

[hint: try 6146384]

Gender: ☒ Male ☐ Female

Ethnicity:

☐ Caucasian

☒ African American

☐ American Indian or Alaska Native

☐ Native Hawaiian or other Pacific Islander

☐ Asian

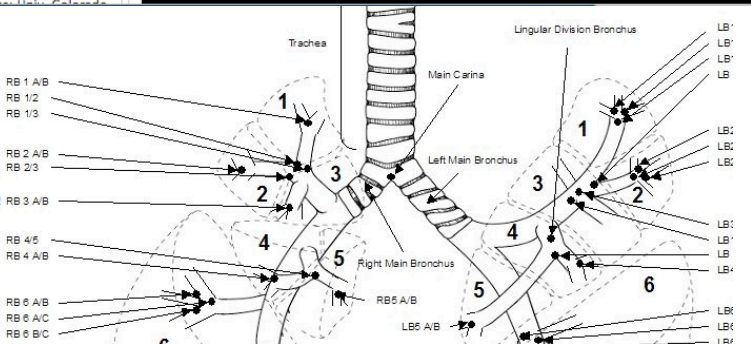
☐ Other

☐ Unknown / refused

Born between and

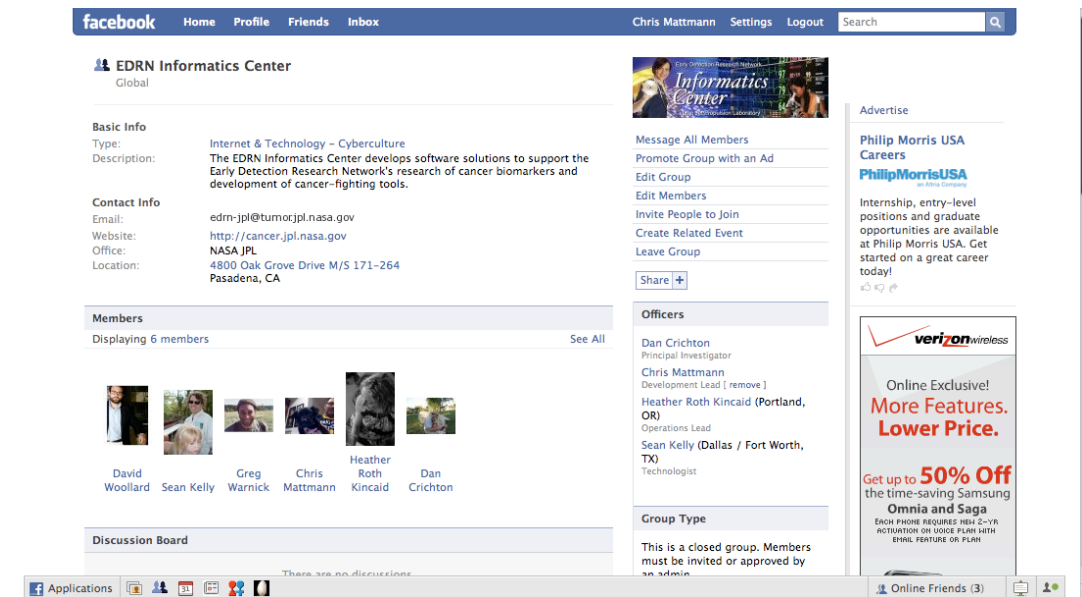
Patient History of Cancer: ☐ No ☒ Yes

Type:



Get involved

- Check out our web presence:
 - <http://cancer.jpl.nasa.gov/>
- Check out our public portal (BETA):
 - <http://edrn.jpl.nasa.gov>
- Check out our facebook group:
 - <http://www.facebook.com/group.php?gid=56938589930>



Acknowledgements

- NCI
 - Don Johnsey
 - Christos Patriotis
 - Sudhir Srivastava
- Data Management and Coordinating Center (DMCC) at Fred Hutchinson Cancer Research Center
- EDRN IC
 - Dan Crichton, Andrew Hart, Steve Hughes, Heather Kincaid, Sean Kelly, John Tran, Thuy Tran

Questions